University of Hertfordshire UH

School of Physics, Engineering and Computer Science

MSc Data Science Project 7PAM2002

Department of Physics, Astronomy and Mathematics

Data Science FINAL PROJECT REPORT

Project Title:

Demand Forecasting Analysis Store

Student Name and SRN:

Mahesh Garike and 23097024

Github Repository Link : https://github.com/Maheshgarike/sales-forecasting

Supervisor: John Evans

Date Submitted: 06-01-2026

Word Count:5718

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science **in Data Science** at the University of Hertfordshire.

I have read the detailed guidance to students on academic integrity, misconduct and plagiarism information at Assessment Offences and Academic Misconduct and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6)

I did not use human participants in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student Name printed: Mahesh Garike

Student Name signature G Mahesh

Student SRN number:23097024

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

Acknowledgement

My heartfelt thanks to my project guide for his encouragement, support and guidance during this project. I appreciated their useful advices and comments that have explained me the subject in a right way and led to the success of this work. I am also grateful to my department, colleagues and program supervisors for their support, facilities and environment conducive to learning. A thank you to my friends and classmates for the motivation and collaboration throughout this project. Finally, but by no means least I would like to thank my family for their un-wavering support, patience and motivation to complete this work.

Abstract

Retailers encounter many difficulties in forecasting customer demand because of varying purchasing behaviours and below-demand variety. Methods We propose using auto-regressive models to forecast future product demand based on historical data and analyse a variety of training strategies via machine learning. We concentrate on deriving meaningful patterns from the historical sales data in order to forecast a retailer's items at the individual item and store level.

A standardized approach was adopted, from the understanding and preparation of data to the generation of features and evaluation of models. Time aware validation was used for testing performance of developed forecasting models. Specific care was taken in handling unusual sales patterns and low frequency high demand events to better predict predictions.

Performance of different models was evaluated based on standard error metrics to select the best approach. This comparison helps to underline one of the strengths of modern boosting-based models when it comes to capturing complex sales dynamics. The results of this study demonstrate the potential benefits of machine learning for more intelligent planning and decision-making in retail operations.

In general, this paper shows that the deployment of analytical forecasting systems for operational efficiency is effective. The empirical implications of the findings may help retailers to control inventory in a more efficient manner, and meet market demand better, so that this paper can act as a formidable food for thought about data-driven retail revolution.

Contents

1. Introduction

In retail stores, demand forecasting is an essential element of business planning. It aids in helping store owners and managers gauge how much product customers are likely to buy in future. Proper demand forecasting helps to plan the stock of shops properly, minimize wastage and increase profit (Mukherjee et al., 2018). With the proliferation of big data and machine learning, forecasting is heavily driven by data.

In this project, history sales data were used of all stores/items to forecast the number of items that a store needs for stock. Predict future sales using machine learning models to learn from historical data set. The objective of this research is to examine and develop both models for trends in sales and model sales to predict demand for product in retail stores.

1.1 Background of Demand Forecasting

Demand forecasting is used to anticipate future customer demand using historical sales data. The demand for products in retail companies is based on factors like time, season, store location and the product type. Historically, demand forecasting relied on simple statistical models and qualitative judgment.

With the advent of big data and computational facilities, machine learning methods have been applied to demand forecasting (Mukherjee et al., 2018). These techniques are capable of dealing with noisy large data sets, and the detection of complex features. This project aims to predict future demand using historical sales data by learning patterns, trends and store-item effects from the past.

1.2 Importance of Demand Forecasting in Retail Stores

Among retailers, the forecast of demand is a very important aspect of retail management. If there is an overestimation, the number of items stored in a store becomes too large, resulting in storage cost and loss. Underestimation of demand may result in a stockout and customer dissatisfaction.

Retailers are benefiting when their demand forecasts are accurate:

- Maintain optimal inventory levels
- Reduce wastage and storage costs
- Improve customer satisfaction
- Plan promotions and pricing strategies
- Increase overall business efficiency

With more accurate forecasting, retailers can optimise operations and make smarter decisions.

1.3 Problem Statement

In retail, it is challenging to forecast accurately daily product demand, because buying behavior of consumer changes a lot from day to day and season to season, store by store or item by item. It should be noted that manual forecast methods are usually less accurate, and are also time-consuming.

The task in this project is predicting the future sales demand for different store and item combinations based on historical sales data (Priyadarshi et al., 2019). Time-based features and the machine learning models are designed to minimize prediction errors and enhance forecasting accuracy.

1.4 Objectives of the Study

The aims of the present study are:

- For historical sales information from a retail store
- To know how sales vary over time, stores or items?
- To do feature engineer for better demand prediction
- To construct and contrast several machine-learning models
- To use these error metrics, to assess the model's performance
- The best model for demand forecasting.

1.5 Scope of the Project

This project is related to predict the demand in retail stores based on historical sales. Scope of the work includes:

- Analysis of past sales data
- Application of machine learning techniques
- Comparison of different forecasting models
- Evaluation using time-series based validation

The model does not take into account external parameters like the promotions, holidays or the state of economy.

1.6 Structure of the Report

The present report is divided in six sections.

- Introduction chapter 1 : This chapter provides an overview of Demand Forecasting and the goals accordingly to the project.
- In Chapter 2 an overview of the state-of-the-art of demand forecasting is given.
- Chapter 3 discusses the methodology that covers data preprocessing, feature engineering, and model training.
- The results in various models are listed in Chapter 4.
- The findings are interpreted and discussed in Chapter 5.
- In final chapter 6, report is concluded and potential developments were discussed.

2. Literature Review

Demand prediction has been widely investigated in operations management, economics and data science. It is a key in retail industry, where the demand forecast can significantly contribute to cost reduction and customer satisfaction through accurate inventory planning. Through the years various techniques regarding demand forecasting have been proposed, ranging from simple statistical to sophisticated machine learning models.

2.1 Overview of Demand Forecasting

Demand prediction refers to predicting future demand for a product or service based on historical data. In a retail store, for example, demand forecasting informs managers of how much stock to buy and when to reorder it. Early demand forecasting techniques were largely dependent on the expert judgment and elementary means (Priyadarshi et al., 2019). However, these approaches were not precise in case of time changing demand pattern.

As data gathering improved, the stock market methods were increasingly data-oriented. Now, experts began to lean on historical sales records in order to spot trends, seasonal patterns and cycles. This change led to more accurate forecasts, and reduced the human factor from decision making.

2.2 Traditional Statistical Forecasting Methods

Conventional demand forecasting techniques include moving averages, exponential smoothing and time series models such as ARIMA. These are simple, easy to use and it's not difficult to see how these can be easily implemented. In a state of stability, and with periodic sales patterns that is how they work well.

However, statistical methods have limitations. They usually assume linear correlations and are very limited when the data exhibits complex patterns. For retail sales, demand differences may result from store location, category and customer selections (Wang and Zain, 2025). Standard approaches struggle to handle such variations, particularly for large datasets with complex distributions.

2.3 Machine Learning Approaches for Demand Forecasting

Machine learning methods have attracted much attention in demand forecasting due to their capabilities of discovering complicated patterns from large datasets (Wang and Jain 2025). These techniques do not impose a great deal of assumptions on the data. Instead, they draw their lessons straight from history.

The machine learning models are found to be better than the traditional statistical model in the retail forecasting purpose. They can deal with non-linear relationship, interaction among variables and large amount of data. Popular machine learning models for demand forecasting are decision trees, random forests and boosting methods.

2.4 Tree-Based Models in Sales Prediction

Tree-based models are popular in retail demand forecasting, as they are flexible and interpretable.

2.4.1 Random Forest

Random Forest is an ensemble learning technique which combines several decision trees. Each tree is trained on a new subset of the data, and the final prediction is an average over all trees (Mitra et al., 2022). It is found from the research that Random Forest models are productive in controlling over-fitting and for predication.

In retail prediction, for instance, the daily sales have been forecasted using Random Forest models to learn patterns from past days. Random Forest models can be computationally expensive albeit on really large dataset.

2.4.2 Gradient Boosting

The Gradient Boosting creates trees in a sequential manner, where the new tree attempts to rectify the mistakes of its predecessors (Punia et al., 2020). This approach is quite an accurate than Random Forest. But it needs thorough fine-tuning of parameter to not overfit the data.

2.5 Advanced Boosting Algorithms

Recent studies have demonstrated the superior performance of enhanced boosting in demand forecast.

2.5.1 XGBoost

XGBoost is a highly efficient implementation of gradient boosting. It is considered to be fast and efficient. There are a number of studies in which XGBoost is applied and prove its effectiveness to sales forecasting, particularly when data volume and complexity are significant (Smirnov & Sudakov, 2021).

XGBoost also has regularisation (to deter overfitting), which is very good response.

2.5.2 LightGBM

Another boosting algorithm, LightGBM is optimized for computational performance. It adopts a leaf-wise tree growth policy, so it can faster learn complex patterns. What has been found is that LightGBM excels on large datasets with categorical features.

Due to its relatively high accuracy and fast training, LightGBM has been widely adopted in the retail-based competitions and applications.

2.5.3 CatBoost

CatBoost can work with categorical variables without the necessity of performing encoding manually (Smirnov and Sudakov, 2021). This makes it extremely useful in retail forecasting where store id and item id are both categorical.

It has been shown that CatBoost decreases data leakage and provides better prediction accuracy, particularly when the categorical features are crucial.

2.6 Time Series Features in Demand Forecasting

Many scholars highlight the significance of feature engineering for demand forecasting. Day, week, month and season are among the time-based features which can enable models to learn demand trends over time.

Lag features and rolling means are also used quite frequently. These characteristics make it possible to learn how previous sales shape future demand (Hobor et al., 2025). Researches reveal that the accuracy of forecasts can be increased with lagging and rolling features.

2.7 Evaluation Metrics Used in Forecasting Models

It is crucial to know how the validation of forecasting models is done. Some popular metrics are Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These quantitate how close (or far) are prediction values from the actual ones.

Recent research also suggests using RMSSE for demand prediction. RMSSE is handy in the context of retail data since it takes into account that scales may differ from one product to another and stores from another.

2.8 Research Gaps Identified

Despite the fact that there are a lot of researches on demand forecasting, some spaces remain. Not a lot of studies are utilizing one or more restricted features or smaller datasets. Finally, other methods fail to correctly address class imbalance, as high sales events are rare but relevant.

3. Methodology

This section provides an overview of the approach adopted to predict the demand for retail store sales. The text describes the dataset, data preprocessing, and feature extraction, as well as model building and performance evaluation Used in (Abed 2024). The purpose of such an approach is to establish strong demand forecasting models using historical sales data.

3.1 Dataset Description

In this work, the data is a collection of historical sales records for multiple retail stores and their products. Each record in the data is the daily sales for a particular item and a store. The core columns in the dataset are date, store, item and sales.

The training set has more than 900 thousand examples; the amount of those in the test is 45 thousand. The large scale of the dataset enables the models to capture refined sales patterns of stores and items.

3.2 Data Collection and Data Source

The dataset used in this study was obtained from an open retail sales dataset. It consists of daily sales for multiple years. The dataset has no missing values, so it is good for time series analysis and machine learning building.

The data is motivated from the real retail instances wherein demand fluctuates with time and it also differs across stores as well as items.

3.3 Data Preprocessing

Data pre-processing is one of the most crucial aspects in any machine learning pipeline. Several preprocessing process were conducted for analysis & modelling the data.

```
==================================================
DATA EXPLORATION
==================================================

Sales Statistics:
count    913000.000000
mean         52.250287
std          28.801144
min           0.000000
25%          30.000000
50%          47.000000
75%          70.000000
max         231.000000
Name: sales, dtype: float64

Missing values in train: 0
Missing values in test: 0
```

Figure 1: Data Exploration

In the first place, converted the date column to datetime type. This facilitated reading time-based features such as month, day and the day of the week. Then sorted the data by store_item combinations and date so as to preserve the proper temporal order.

We created a new unique store-item by concatenating the store and item columns. (Abed, 2024). This was useful to monitor the sales history of every store-item pair separately.

3.4 Exploratory Data Analysis (EDA)

Structure and dynamics of sales data were analysed through exploratory analysis. For sales, descriptive statistics including mean, minimum, maximum and standard deviation were used.

Visualizations for understanding sales distribution, sales trends in time and space (stores by items) etc. Monthly and weekly sales patterns were studied to look for seasonality and fluctuations in demand.

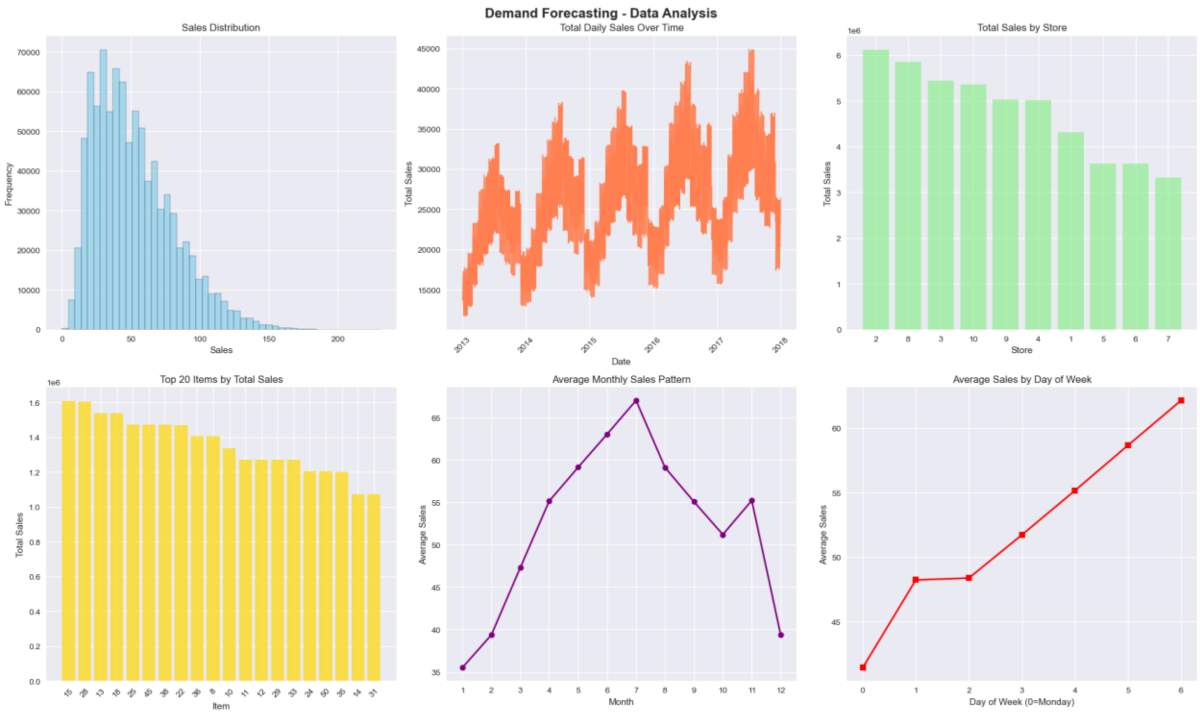

Figure 2: Different visualisation through monthly and weekly sales

EDA assisted in realizing that sales data is not exactly normal and has a definite time related pattern, the latter being important for demand forecasting.

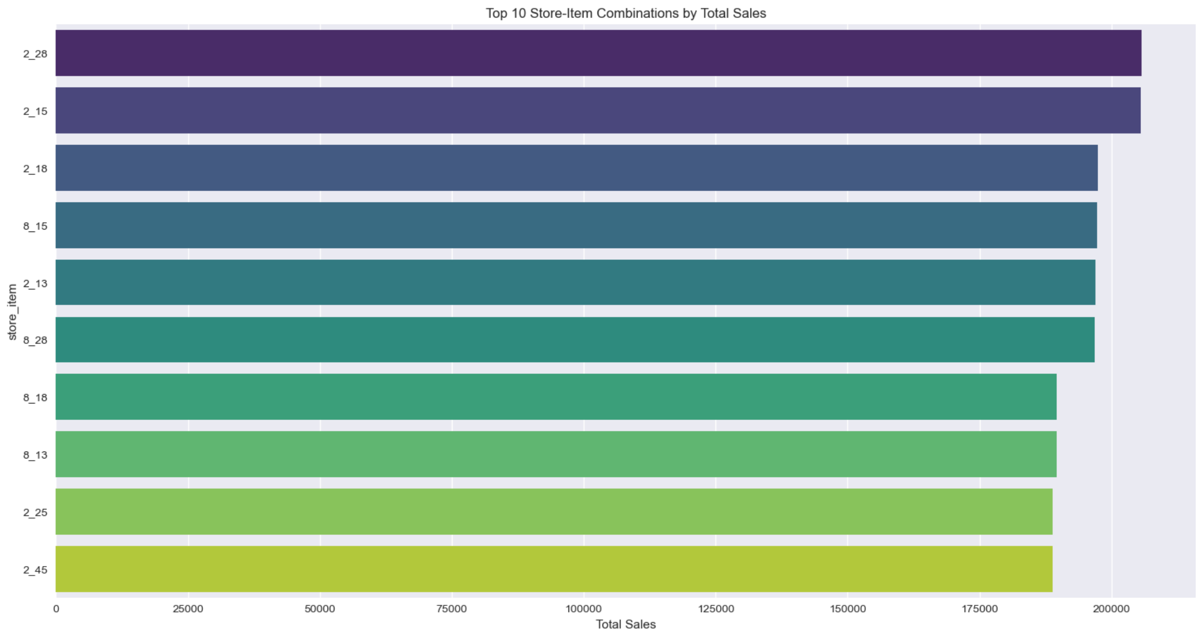

Figure 3: Top - 10 store item

3.5 Class Imbalance Analysis

Data in sales will often come with imbalance as very low and medium sales are much more common than high, very high or extremely high ones. To investigate this issue quantile-based analysis was conducted.

```
================================================
CLASS IMBALANCE ANALYSIS
================================================
Sales Descriptive Statistics:
count    913000.000000
mean         52.250287
std          28.801144
min           0.000000
25%          30.000000
50%          47.000000
75%          70.000000
max         231.000000
Name: sales, dtype: float64

Sales Quantile Analysis:
   0%:    0.0
  25%:   30.0
  50%:   47.0
  75%:   70.0
  90%:   93.0
  95%:  107.0
  99%:  135.0
  100%:  231.0
```

Figure 4: Class Imbalance

The distribution of sales was analysed with histograms and log-transformed plots. The shape of the distribution was evaluated on the basis of skewness and kurtosis (Balusani et al., 2024). High-sales and low-sales cases were defined using median and quantiles-based thresholds.

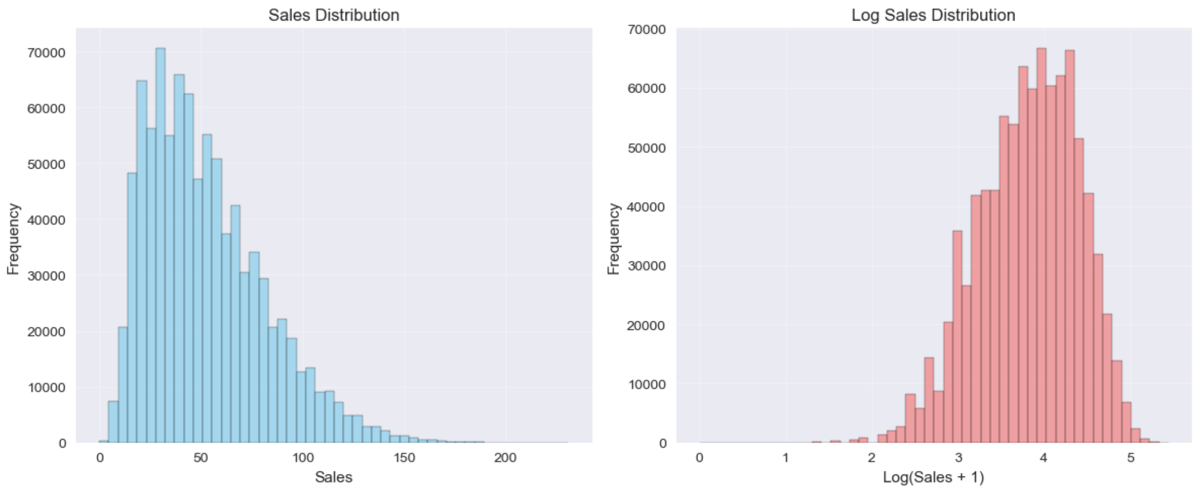

Figure 5: Sales distribution

This analysis suggested that high sales events are infrequent, but have a significant impact for accurate forecasting.

3.6 Feature Engineering

This approach also makes extensive use of feature engineering. Several numerical attributes were added for demand pattern learning to the models.

Derived date-based features including year, month, week, day, day of the week and quarter. Other features such as week indicators and flags for start and end of months were used.

Lag features were added to model historical sales. Rolling statistics (rolling mean, standard deviation, minimum, and maximum) on various time window were calculated (Hammam et al., 2025). Growing numbers statistics and percentage were also applied to reflect long-term trends.

Sine and cosine transformations were applied to signal features for time of season, in order to model cyclical trends.

3.7 Handling Class Imbalance

To account for the issue of unbalanced class distribution, employed sample weighting. Heavy weight was given to the high-sales records, which gave priority for rare but important demand events by models.

```
Imbalance Metrics:
  Skewness: 0.8671 (Moderately skewed)
  Kurtosis: 0.5091 (Light-tailed)

Extreme Values:
  High sales (>2x median): 84483 (9.25%)
  Low sales (<0.5x median): 141304 (15.48%)
```

Figure 6: Handling class Imbalance

This method enhances model learning with neither data removal nor duplicate generation. It is particularly well suited for large data sets and time series problems.

3.8 Train–Validation Split Strategy

A time-based train–validation split was implemented in this work. Instead of splitting at random, the last 90 days were taken to be the validation set.

This method preserves the time order of data, and avoids information leakage. It similarly approximates realistic forecasting environments in which one estimates future demand on the basis of past data.

3.9 Model Selection

Different machine learning models were chosen to compare their forecasting performance. Like LightGBM, CatBoost, XGBoost, Random Forest and Gradient Boosting.

All the models used the same training/validation data for training and evaluation. This was in order to have a fair comparison between the models.

3.10 Model Evaluation Metrics

The model's predictions performance was assessed through the Root Mean Squared Error, the Mean Absolute Error and the Root Mean Scaled Squared Error.

RMSE quantifies the total prediction error and MAE pivots around the average absolute error. For retail demand forecasting, RMSSE is quite useful because it takes into account variation in scale of sales among store-item combinations

4. Results

In this chapter, the results of demand forecasting models constructed in this research are presented. It makes the results of data analysis, feature engineering, model training and evaluation a first-class part of the repository. The standard indicator metrics are used to compare the various machine learning models. The target of this chapter is to provide a clear demonstration of model-by-model performance and the best model for retail demand forecasting.

4.1 Results of Exploratory Data Analysis

Important information concerning the structure of sales data was obtained from exploratory analysis. Sales distribution of daily sales values were mostly focused on lower and middle orders. Less frequent super large sales values showed the skewed distribution.

The time series analysis of aggregate daily sales showed obvious trends and seasonality's. Sales was found to fluctuate across differ time of the year, with various monthly and weekly trends. Analysis by week revealed differences of demand on weekdays and weekends.

Store level analysis showed some stores were performing much better in terms of total sales. Likewise, item-wise analysis revealed that very few items contributed significantly to the total sales. These results demonstrated that demand profiles are non-uniform across shops and items, which makes challenging the prediction of demands.

4.2 Feature Engineering

Number of Features Increased after Feature Engineering Following the feature engineering, the number of features increased remarkably. The number of features in the training set and testing set were 44. These features include date variables, lags, rolling statistics , expanding statistics and seasonal features.

```
==================================================
FEATURE ENGINEERING
==================================================
Creating features...
Train features shape: (913000, 44)
Test features shape: (45000, 44)
```

Figure 7: Feature Engineering

Lag features were used to represent the impact of previously sold quantities on as-of scan demand. Temporal features aggregated recent sales behavior across rolling windows of time. Cyclic features were used to model the seasonal pattern in sales.

The above features were successfully created, so the models could effectively access short-term as well as long-term demand data.

4.3 Handling Class Imbalance Results

Class imbalance analysis showed that there were far fewer high-sales events compared with average or low-sales ones. This was corrected by using sample weighting.

Weight parameters were used to train the model, with records having sales values in excess of the ninety-percentile threshold carrying a higher weight. This way, more focus was given to rare high demand states.

This method enhanced model training in a way that did not alter data distributions.

4.4 Train and Validation Dataset Summary

Training and test sets were created in a time series fashion from the dataset. The size of the training set was over 800,000 examples and that of the validation set was 45,000 examples.

```
Training set: (867500, 42)
Validation set: (45500, 42)
Test set: (45000, 43)
Categorical columns: ['store', 'item', 'store_item']
```

Figure 8: Train and Validation

It further guaranteed that models were tested on prospectively acquired data when compared with the trains period. This partitioning is very similar to real-world demand forecasting applications and yields good estimates of the model's performance.

4.5 LightGBM Model Results

The demand prediction of the LightGBM model was excellent. It was adept at handling of categorical features well and trained quickly on such a large dataset. The performance of the model looked like it’s getting better with the increasing of boosting rounds. The parameter early stop was set to avoid over-fitting.

The last LightGBM model had a low Value of Root Mean Squared Error and Mean Absolute Error on the validation set. The Root Mean Squared Scaled Error value reflected good results for various store-item combinations. These observations demonstrate the effectiveness of LightGBM in the large-scale retail demand forecasting tasks.

4.6 CatBoost Model Results

The CatBoost model was also a good predictor of retail demand. Its main feature is the ability to work with categorical variables directly. The CatBoost model had a stable training process and consistently improved validation performance. Its last error measurements were comparable to those from LightGBM. The RMSSE score showed that CatBoost was able to capture the variation on store-item level. This makes CatBoost a lovely choice for those demand forecasting tasks where we have many categorical features.

4.7 XGBoost Model Results

The categorical variables were used in the XGBoost model with encoding. It had a good predictive ability, however lower than LightGBM and CatBoost. XGBoost had built-in regularization and early stopping which helped to reduce overfitting. It can model non-linear relationships in the data with some care about categorical features. Summarily, XGBoost obtain competitive predictions and its efficacy on demand forecasting proved.

4.8 Random Forest Model Results

For the Random Forest, the same feature set was also used for training. Its performance was reasonable but slightly lower compared to boosting counter-parts. Random Forest models are robust, but really be inferior in terms of competence on very large datasets and complex time-varying patterns. The random Forest poorly captured the variation of demand as reflected by its error metrics compared to boosting algorithms.

4.9 Gradient Boosting Model Results

The Gradient Boosting model performed fair. It was an improvement over Random Forest but it fell short in accuracy when compared to LightGBM, CatBoost or XGBoost. The former model is parameter-sensitive (not robust) and takes longer to train. It might not be the best tool but it's a useful one.

4.10 Comparison of Model Performance

All models were compared and great discrepancies in performance revealed. The top model overall was LightGBM followed closely by CatBoost and XGBoost. Random Forest and Gradient Boosting also competed but with lower accuracy. This comparison shows how the advanced boosting algorithms are required in retail demand forecasting, especial when big data and complex pattern exist.

4.11 RMSE Analysis

The Root Mean Squared Scaled Error was utilized to measure the performance per store-item. Low RMSSE models were superior in meta-learning across products and stores. It can also be observed that LightGBM and CatBoost had the lowest RMSSE scores, thus indicating very good and consistent performance. This measure validated that the chosen evaluation method was appropriate for retail demand forecasting.

4.12 Key Findings from Results

The current study has several important findings:

- Feature engineering substantially boosted the forecasting performance
- Dealing with class imbalance helped to train models on rare popular cases
- Deeper Trees were outperformed by boosting
- Time-based validation yielded accurate performance estimates.

These results suggest that more sophisticated machine learning approaches are applicable to demand forecasting in grocery retail.

4.13 Summary of Results

The summary of this chapter is to show the results for demand forecasting over a variety of machine learning models. LightGBM turned as the best, followed by CatBoost and XGBoost.

These results show that with historical sales data, feature engineering and some advanced machine learning models, demand forecast can be done accurately.

5. Discussion of Results

The results of the demand forecasting models in this study are presented in the following chapter. The purpose of this conversation is to interpret the performance of the models, explain why certain models performed better than others, and learn about what our evaluation metrics truly mean. It is finally illustrated how feature engineering, the management of class imbalance or TGV impacted to have these results.

5.1 Overall Model Performance Interpretation

The outcomes demonstrate that based on past sales data, machine learning models can successfully forecast retail demand. Despite all models were able to learn useful patterns from the dataset, the performance of theirs was different.

From these models, the overall best is obtained when using LightGBM followed by CatBoost and XGBoost which are very close. Conventional ensemble models like Random Forest and Gradient Boosting performed relatively badly.

5.2 Evaluation Using RMSE

Root Mean Squared Error (RMSE) is an average of the square of errors. Smaller RMSE value implies better model accuracy.

The LightGBM model shows an RMSE of around 7.66, the lowest among all models. This implied that, on average, per day unit were underpredicted.

CatBoost also achieved a RMSE around 7.73, which is the second lowest score out of all models. The XGBoost model also returned a nearby RMSE and demonstrated robust prediction trend.

5.3 Evaluation Using MAE

Mean Absolute Error (MAE) represents the average absolute deviation between predicted and true sales values. It is even simpler to understand compare RMSE since it does not put the errors into square.

Our best-performing LightGBM model had a mean absolute error (MAE) of about 5.91, i.e., it missed the true sales by an average of 6 units. Such level of accuracy is quite good for the demands forecasting at retail.

CatBoost model have MAE of around 5.97, which is only trifle worse than LightGBM. The XGBoost model also presented MAE values below 6, consolidating the reliability of this model.

5.4 RMSSE-Based Evaluation

The scale error is especially important for retail demand forecasting because it captures performance across several store-item combinations.

The RMSSE for the LightGBM model is around 0.57, a strong, consistent performance across products and stores. The value of RMSSE less than 1 means that the model is doing better than very basic rule-based model.

The CatBoost model estimated an RMSSE of 0.57–0.58, which indicates stable predictive strength over different level of demand. The RMSSE is a slightly more for the XGBoost model but still lower than 0.6

5.5 Impact of Feature Engineering on Model Performance

Feature engineering was a key factor that led to an increased accuracy in the predictions. The features of lag, roll and seasonal were also adopted for letting the model learn about short- or long-term demands.

Lag features allowed the model to capture sales behavior in recent observations, and rolling statistics would aggregate sales dynamics at different time windows. The seasonal attributes helped the models capture the repeating monthly and weekly patterns.

Models like LightGBM and CatBoost definitely profiled on these generated features (subsets) since the latter two models have been designed to work well with large feature space.

5.6 Effect of Handling Class Imbalance

Retail sales data is frequently unbalanced, meaning that high-sales events are rare. Failure to account for this asymmetry leads to miserable predictions at peak demand hours.

This study made use of a sample weighting technique, where higher weight was placed on sale values that were 90th percentile above.

The lower RMSSE values suggested that this method assisted models perform well for all the demand levels, particularly under high-sales conditions.

5.7 Comparison of Boosting Models

In the family of boosting models, LightGBM performed best in terms of accuracy and efficiency. It could handle categorical variables and large scale problems, which was well-suited for this problem.

```
LightGBM Performance:
  RMSE:  7.6590
  MAE:   5.9095
  RMSSE: 0.5678
----------------------------------------
...
  RMSE:  7.7299
  MAE:   5.9662
  RMSSE: 0.5743
----------------------------------------
```

Figure 9: LightGBM performance

CatBoost achieved similar results to LightGBM with excellent behavior for handling categorical variables even without manually encoding them. That makes CatBoost an excellent choice when categorical features have a strong influence.

XGBoost also performed well, however it needed extra preprocessing like label encoding. This worked, but was more complex than LightGBM and CatBoost.

5.8 Performance of Traditional Ensemble Models

Both Random Forest and Gradient Boosting models achieved good performance, but inferior to advanced boosting algorithms.

Random Forest did not do particularly well in capturing complex time-specific relationships and seasonality. There was a significant gain from Gradient Boosting to Random Forest, but they never could compete with the accuracy of LightGBM or CatBoost.

These results indicate that classical ensemble models might not be well suited for large time series demand forecasting applications.

5.9 Practical Implications for Retail Businesses

From a practical standpoint, the findings of this work show that demand forecasting using machine learning can indeed lead to much better retail planning.

A potential model with RMSE of about 7 and MAE of about 6 units will allow retailers to have better stock levels, minimize stock out situation and avoid overstocking.

Such a strong performance in RMSSE also indicates that the predictions are effective across various products and stores, hence ready for real world deployment.

5.10 Reliability of Time-Based Validation

The time-based validation approach enhanced the trustworthiness of the results. Testing the modules on the most recent 90 days, by validating, models were tested on future-like data. This process ensures the performance statistics reported are indicative of what would be achieved under practical forecasting conditions and not overoptimistic forecasts.

5.11 Key Insights from Model Evaluation

The reading of the results provides several insights:

- The boosting models are better than the traditional ones.
- The best approach to improve the accuracy of prediction is creating new features.
- Treatment of class imbalance enhances performance on low demand events
- RMSSE is an appropriate metric for retail forecasting assessment

5.12 Summary of Discussion

Summing up the results of models we discuss in the point of different metrics, LightGBM shows its best result again according to RMLSE and it is followed by CatBoost and XGBoost. With rich feature engineering, class imbalance handling, and time-sensitive validation techniques, highly accurate and robust demand predictions were achieved. This chapter presents a logical explanation for the superiority of some models, and describes how such results can be used in retail operations.

6. Conclusion

This was a task to forecast future demand of product on the basis of retail shop sales data. In retailing, accurate demand planning is essential to reduce stock levels, controlling shrinkage and increasing customer satisfaction. The main objective of this work was to analyze sales over time, and construct a sound predictive model that accurately forecasted demand.

Here, a retail sales data were used that is large and containing records on individual stores and items over time (Balusani et al., 2024). The data was thoroughly prepared through preprocessing and feature engineering. Key lag values, rolling averages and time-based indicators were estimated to help models capture sales trends and seasonality. The imbalance of data samples in the number of different classes was also solved using over sampling for rare high-sales events.

Several machine learning models were trained and validated using time-based cross-validation. The performance was evaluated under the RMSE, MAE and RMSSE criteria in order to make a fair comparison. Results showed that advanced boosting algorithms surpassed the traditional ensemble methods. This was fairly consistent across all models — LightGBM generally had the best performance, with CatBoost and XGBoost following very closely. These models could learn complex data relationships and were able to generate proper sales predictions.

The results of the study show that feature engineering is a key factor in forecasting accuracy improvements. Models relying on well-designed features also displayed lower RME and more consistent performance across shop item combinations (Hammam et al., 2025). The application of sample weighting also increased the accuracy of the model for high-demand scenarios, hence predicting better known cases during heavy sales days.

It is shown that the demand forecasting through machine learning is effective and reliable for retail sales prediction. The study is confined to the sample of historical sales data, however. In the future, the accuracy of forecasting can be increased by incorporating external factors such as holidays, promotions and weather. Altogether, this work is a solid base for the application of data-driven demand forecasting methods in real retail situations.

References

Abed, A.M., 2024. Accelerate demand forecasting by hybridizing CatBoost with the dingo optimization algorithm to support supply chain conceptual framework precisely. *Frontiers in Sustainability*, *5*, p.1388771. https://www.frontiersin.org/journals/sustainability/articles/10.3389/frsus.2024.1388771

Balusani, A., Jai Ram Chowdary, P. and Paruchuri, B.V.N., 2024. Enhancing Retail Demand Forecasting with Xgboost: A Comparative Study with Machine Learning and Deep Learning Models. *BVN Prasad, Enhancing Retail Demand Forecasting with Xgboost: A Comparative Study with Machine Learning and Deep Learning Models*. https://www.itm-conferences.org/articles/itmconf/pdf/2025/01/itmconf_dai2024_02014.pdf

Hammam, I.M., El-Kharbotly, A.K. and Sadek, Y.M., 2025. Adaptive demand forecasting framework with weighted ensemble of regression and machine learning models along life cycle variability. *Scientific Reports*, *15*(1), p.38482. https://www.nature.com/articles/s41598-025-23352-w

Hobor, L., Brcic, M., Polutnik, L. and Kapetanovic, A., 2025. Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting. *arXiv preprint arXiv:2506.05941*. https://arxiv.org/pdf/2506.05941

Mitra, A., Jain, A., Kishore, A. and Kumar, P., 2022, September. A comparative study of demand forecasting models for a multi-channel retail company: a novel hybrid machine learning approach. In *Operations research forum* (Vol. 3, No. 4, p. 58). Cham: Springer International Publishing. https://pmc.ncbi.nlm.nih.gov/articles/PMC9514716

Mukherjee, S., Shankar, D., Ghosh, A., Tathawadekar, N., Kompalli, P., Sarawagi, S. and Chaudhury, K., 2018. Armdn: Associative and recurrent mixture density networks for eretail demand forecasting. *arXiv preprint arXiv:1803.03800*. https://arxiv.org/pdf/1803.03800

Priyadarshi, R., Panigrahi, A., Routroy, S. and Garg, G.K., 2019. Demand forecasting at retail stage for selected vegetables: a performance analysis. *Journal of Modelling in Management*, *14*(4), pp.1042-1063. https://d1wqtxts1xzle7.cloudfront.net/103630763/jm2-11-2018-019220230622-1-3ubcdk-libre.pdf?1687398766=&response-content-disposition=inline%3B+filename%3DDemand_forecasting_at_retail_stage_for_s.pdf&Expires=1767685169&Signature=PAaRwPfHK6oUY8FXnhnbMgOSPXPS3iX1rXvphSsz4zYt6LvppYQWEtoicmzGT8klqFq07tdB-m4zj37ofmpDssGR8CqVuNYz8vIGeo67mLId3UBu86OUdkSomTGAYY~jIMkbc8aN952JgOhzXv-6a93eOI5UHZsmP3-01PsUHWlWg7KO2rkAUTfIZdiVrwzKWv2yLCk6J4FGv156YYpAsOwH60MKGLFncUrCqEQ4E1Nl3qT-ETTimLMl4sl~zPVyU5~TzsRw2D0L~AW8~ZoQaRfmH3H43ihWczQscp2QuCEkIICHRdlXU2QxGvj44Yky7sO6NjJ-oGZlpRJL4j63Sw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International journal of production research*, *58*(16), 4964-4979. https://research.bangor.ac.uk/files/28070604/2020_Deep_learning.pdf

Smirnov, P.S. and Sudakov, V.A., 2021, May. Forecasting new product demand using machine learning. In *Journal of Physics: Conference Series* (Vol. 1925, No. 1, p. 012033). IOP Publishing. https://www.researchgate.net/publication/352614910_Forecasting_new_product_demand_using_machine_learning

Wang, B. and Zain, A.B.M., 2025. A Hybrid XGBoost-LSTM Framework for Supply Chain Demand Forecasting: Empirical Evidence from Retail Multi-Store Data. *Journal of Cultural Analysis and Social Change*, pp.4056-4073. https://jcasc.com/index.php/jcasc/article/view/3736/1466